

# Reasoning patterns in autism: rules and exceptions\*

Michiel van Lambalgen<sup>†</sup>      Heleen Smid

June 14, 2004

## 1 Introduction: psychiatric disorders and reasoning

Autism is a clinical syndrome first described by Leo Kanner in 1943, typically first diagnosed in early infancy. Patients are often referred because of delayed language onset, deficits in affective relations and communication (such as refusal of eye contact), and quite often also abnormal motor behaviour, such as repetitive movements (often self-harming) and indifference to resulting pain. Autistic children characteristically do not engage in spontaneous phantasy play, although they may be able to phantasize when instructed to do so. They are quite literal-minded, and do not understand metaphorical use of language. Today, autism is often referred to as ‘Autism Spectrum Disorder’ (ASD), which reflects the fact that the disorder occurs in all severities – from a complete lack of language and severe retardation to a ‘normal’ range of IQs and level of functioning.

More than any other psychiatric disorder, autism has captured the imagination of the practitioners of cognitive science, because, at least according to some theories, it holds the promise of revealing the essence of what makes us human. This holds especially for the school which views autism as a deficit in ‘theory of mind’, the ability to represent someone else’s feelings and beliefs. Some go so far as to claim that in this respect autists are like our evolutionary ancestors, given that chimpanzees have much less ‘theory of mind’ than humans. Although we believe such claims need to be qualified, we still agree that autism is important from the point of view of cognitive science. It should perhaps be noted that considering psychiatric disorders cognitively entails

---

\*To appear in Luis A. Perez Miranda and Jesus M. Larrazabal (eds.), *Proc. International Colloquium on Cognitive Science* Donostia/San Sebastian 2003 (Kluwer)

an information-processing view of human beings, an ‘objectivizing’ view which some may find inappropriate in this context, although on a sufficiently broad conception of ‘information’, also emotional processes are informational.

In this article we present some experimental results on propositional reasoning by autists, and to motivate our approach we will start with some general remarks on the relation between (the study of) reasoning and (the study of) cognition. How human beings reason is studied in a subfield called ‘psychology of reasoning’. Often one finds deviations from classical logic, which is generally (and unfortunately) taken as the norm. But what is most interesting in these so-called deviations is what they may tell us about cognitive functions which subserve reasoning processes, such as long-term memory, working memory, language processing or spatial thinking. A few examples may make this clear. Suppose we define an argument to be *valid* if the conclusion is true whenever the premisses are true. In this sense the following two arguments are valid

All bears in the North are white.  
Novya Zemlaya is in the North.  
∴ The bears on Novya Zemlaya are white.

Not all plops are gaga.  
All bleebbs are gaga.  
∴ Some plops are not bleebbs.

Logicians consider these arguments to be valid but subjects in experiments may think differently. When in the 1930’s the Russian psychologist Luria put the first argument to illiterate peasants in Kazachstan, a typical answer was: ‘How can I know, I’ve never been to the North!’ This type of answer is in fact common among nonliterate subjects, as verified for instance by Scribner in Liberia in the 1970’s, and it may tell us something about cognition, in this case for instance that subordinating one’s personal experience to what the experimenter tells one to assume (just the two premisses) is very hard to achieve.

In the second example even subjects with high levels of education may say that actually nothing follows from the premisses. One reason for this could be that computation leading from premisses to conclusion overloads working memory. There are two ways to arrive at the conclusion: (1) starting from the assumption that all plops are bleebbs and reasoning to a contradiction (‘if all plops are bleebbs and all bleebbs are gaga, then all plops are gaga, quod non’), or (2) using a diagram in which circles represent the extent of the predicates. Both ways comprise nontrivial computations; for instance in case (2) the relative positions of

the circles are not fixed by the premisses, which therefore necessitates a check whether the configurations are as general as can be.

Another reason why subjects refuse to draw a conclusion in this case could be that they believe that, pragmatically, a statement of the form ‘Some  $X$  is not  $Y$ ’ is so uninformative as to be useless in communication (cf. Chater and Oaksford [2]).

Reasoning can thus not be viewed in isolation from other cognitive processes, and, concomitantly, reasoning can be used to gain information about these processes. We will now see how this works out in the case of cognitive functions implicated in autism.

## 2 Empathy, ‘theory of mind’, and reasoning

A famous experiment, the ‘false belief’ task, investigates how autistic subjects reason about other people’s belief. It is thus concerned with a ‘cognitive’ version of empathy – how this is related to empathy with respect to emotions unfortunately cannot be explained here. The standard design of the experiment is as follows. A child and a doll are in a room together with the experimenter. The doll and child witness a bar of chocolate being placed in a box. Then the doll is brought out of the room. The child sees the experimenter move the chocolate from the box to a drawer. The doll is brought back in. The experimenter asks the child: ‘Where will the doll look for the chocolate?’ The answers to this question reveal an interesting cutoff point, and a difference between autists and normally developing children. Before about 3.5 yrs, the normally developing child responds where the child knows the chocolate to be (i.e. the drawer); after 3.5 yrs, the child responds where the doll must falsely believe chocolate to be (i.e. the box). By contrast, autists go on answering ‘in the drawer’ for a long time.

This experiment has been repeated many times, in many variations, with fairly robust results. Some versions can easily be done at home. There is for instance the ‘Smarties’ task, which goes as follows. Unbeknownst to the child-subject, a box of Smarties (also known as ‘M&M’s’) is emptied and refilled with pencils. The child is asked: ‘What do you think is in the box?’, and it happily answers: ‘Smarties!’ It is then shown the contents of the box. The pencils are put back into the box, and the child is now asked: ‘What do you think your [absent] mother will say is in the box?’ We may then observe the same critical age: before age 3.5, the child answers: ‘Pencils!’, whereas after age 3.5 the child will say: ‘Smarties!’

Another version<sup>1</sup> uses an episode from the ‘Bob the Builder’ children’s television series, in which Bob climbs a ladder to do some repair work

on the roof of a house. While Bob is happily hammering, the series' resident gremlin Naughty Spud takes away the ladder to steal apples from a nearby apple tree. After Bob has finished his work on the roof, he makes preparations to climb down. At this point the video is stopped and the child who has been watching this episode is asked: 'Where does Bob think that the ladder is?' Again, children below the cutoff age answer: 'At the tree'.

The outcomes of these experiments have been argued to support the 'theory of mind deficit' hypothesis on the cause of autism. Proposed by Leslie in 1987, it holds that human beings have evolved a special 'module' devoted specifically to reasoning about other people's minds. As such, this module would provide a cognitive underpinning for empathy. In normals the module would constitute the difference between humans and their ancestors – indeed, chimpanzees seem to be able to do much less in the way of mind-reading. In autists, this module would be delayed or impaired, thus explaining abnormalities in communication and also in the acquisition of language, if it is indeed true that the development of joint attention is crucial to language learning (as claimed for instance by Tomasello [16]).

This seems a very elegant explanation for an intractable phenomenon, and it has justly captured the public imagination. Upon closer examination the question arises whether it is really an explanation rather than a description of one class of symptoms. For instance, the notion of a 'module' is notoriously hazy. In this context it is obviously meant to be a piece of dedicated neural circuitry. In this way, it can do the double duty of differentiating us from our ancestors and being capable of being damaged in isolation. But it is precisely this isolation, or 'encapsulation' as Fodor called it, that is doubtful. 'Theory of mind' requires language to formulate beliefs in<sup>2</sup>and it also entails a considerable involvement of working memory, as can be seen in 'nested' forms of theory of mind, as in 'Shakespeare intended us to realize that Othello believes that Iago knows that Desdemona is in love with Cassio'. However, as soon as one realizes that a 'module' never operates in isolation, then the 'theory of mind deficit' hypothesis begins to lose its hold. We are now invited to look at the (possibly defective) interactions of the 'module' with other cognitive functions (language, working memory, ...), which leads to the possibility that defects in these functions may play a role in autism. And there is of course also the problem of what the 'module' would have to contain, given that for instance reasoning about other people's desires is possible for both autists and nonhuman primates.

### 3 Executive disorder and the box task

In the following we shall look at a rival explanation of autism as *executive dysfunction*, that is, as a failure of executive function. A general definition of executive function is ‘the appropriate initiation and inhibition of actions’; perseveration is viewed here as a dysfunction of inhibition. This explanation assumes that autism is caused by some form of frontal brain abnormality leading to perseverative behaviour, and the inability to switch between tasks spontaneously, even when the context requires this. An interesting comparison with other frontal lobe abnormalities is provided by Melges [9], who discusses patients with acquired frontal lobe lesions. It is believed that such lesions may interfere with the action-selection capacity of working memory. Patients with frontal lobe lesions may become a slave to the demand characteristics of the present. Melges cites two examples: one patient who was shown a bed with the sheet turned down, immediately undressed and got into bed; another patient, who was shown a tongue depressor, took it and proceeded to examine the doctor’s mouth. What is striking about these examples is that patients become dependent on the Gibsonian ‘affordances’ of their environment, which then act almost like stimulus–response bonds. Affordances (as defined by Gibson [4]) are the functional roles that objects may ‘wear on their sleeves’: in this sense a door ‘affords’ to go through it, and a bed with the sheet turned down ‘affords’ to go to sleep in it. But healthy humans use an affordance as a possibility only, to be used in the selection of actions toward a goal, and not as a necessity, i.e. as a condition–action rule. Indeed, ‘deficient context processing’ seems to be an important aspect of executive dysfunction that is amenable to a logical analysis.

To investigate this, we used a reasoning paradigm called the ‘suppression task’, which was identified by one of us ( in [13]) as testing sensitivity to context and the capacity for flexible planning. Details will be added later, but for now it suffices to say that the connection between this particular reasoning task and autism is this: executive function is called upon when a plan has to be redesigned by the occurrence of unexpected events which make the original plan infeasible. Autists indeed tend to suffer from rather inflexible planning. A particularly strong form of this is the inability to inhibit the pre-potent response to a stimulus, even when it is known that the response is inappropriate. This phenomenon is illustrated in an experiment designed by Hughes and Russell [5], the ‘box task’ which as will be seen later lends itself particularly well to a logical analysis.

The task is to get the marble which is lying on the platform (the

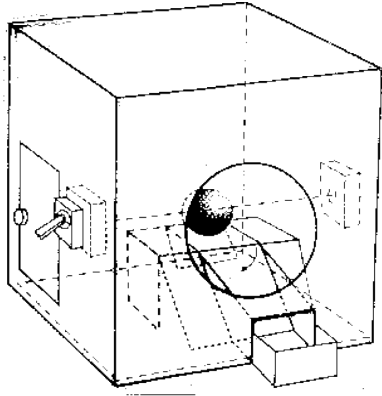


Figure 1: Russell's box task

truncated pyramid) inside the box. However, when the subject puts her hand through the opening, a trapdoor in the platform opens and the marble drops out of reach. This is because there is an infrared light-beam behind the opening, which, when interrupted, activates the trapdoor-mechanism. The switch on the left side of the box deactivates the whole mechanism, so that to get the marble you have to flip the switch first. In the standard setup, the subject is shown how manipulating the switch allows one to retrieve the marble after she has first been tripped up by the trapdoor mechanism.

The results show a striking similarity to the phenomena exhibited in the false belief task: normally developing children master this task by about age 3.5, and before this age they keep reaching for the marble, even when the marble drops out of reach all the time. Autistic children go on failing this task for a long time. The performance on this task is conceptualized as follows. The natural, 'pre-potent', plan is to reach directly for the marble, but this plan fails. The child then has to re-plan, taking into account the information about the switch. After age 3.5 the normally developing child can indeed integrate this information, that is, inhibit the pre-potent response and come up with a new plan. It is hypothesized that autists cannot inhibit this pre-potent response because of a failure in executive function. But to add precision to this diagnosis we have to dig deeper.

It is important to note here that the ability to plan and re-plan when the need arises due to changed context, is fundamental to human cognition, no less fundamental than 'theory of mind' abilities. Human beings act, not on the basis of stimulus-response chains, but on the basis of (possibly distant) goals which they have set themselves.footnoteActually

both humans and nonhuman primates engage in planning. Primates are adept at planning, as has been known since Köhler's 1925 observations [6]. It has even been attested in monkeys. In recent experiments with squirrel monkeys by McGonigle, Chalmers and Dickinson [8], a monkey has to touch all shapes appearing on a computer screen, where the shapes are reshuffled randomly after each trial. The shapes come in different colours, and the interesting fact is that, after extensive training, the monkey comes up with the plan of touching all shapes of a particular colour, and doing this for each colour. This example clearly shows the hierarchical nature of planning: a goal is to be achieved by means of actions which are themselves composed of actions. Consideration of planning therefore provides a measure of evolutionary continuity, while also explaining discontinuity, due to the possibility for goals to become ever more distant. That goal, together with a world-model lead to a plan which suffices to reach the goal in the assumed circumstances. But it is impossible to enumerate *a priori* all events which might possibly form an obstacle in reaching the goal. This is an instance of what is known as the 'qualification problem' in Artificial Intelligence: one can never enumerate all the possible preconditions of an action – and therefore it is generally wise to keep open the possibility that one has overlooked a precondition, while at the same time not allowing this uncertainty to inhibit one's actions. It is possible that at the moment we are writing this, someone publishes a paper outlining the very same ideas, thus making this one otiose – but it would be silly to incorporate this possibility into our present plans for finishing the paper. Nevertheless, we must be prepared to replan when this eventuality arises; and it is perhaps this flexibility that autists are lacking. Indeed, Russell writes (following an unpublished suggestion of Donald Peterson)

[T]aking what one might call a 'defeasibility stance' towards rules is an innate human endowment – and thus one that might be innately lacking . . . [H]umans appear to possess a capacity – whatever that is – for abandoning one relatively entrenched rule for some novel ad hoc procedure. The claim can be made, therefore, that this capacity is lacking in autism, and it is this that gives rise to failures on 'frontal' tasks – not to mention the behavioral rigidity that individuals with the disorder show outside the laboratory [12, p. 318].

Russell goes on to say that one way this theory might be tested is through the implication that "children with autism will fail to perform on tasks which require an appreciation of the defeasibility of rules such as 'sparrows can fly'." This is a sensible suggestion, but we first require a logical description of the box task in order to find a corresponding task in the domain of verbal rules.

### 3.1 The logic of the box task

At first glance there does not appear to be much logic involved in the box task. At a coarse level of formalization we are just concerned with two rules: the first rule applied is

If you put your hand through the opening, you can retrieve the marble.

which is then replaced by

If you throw the switch and put your hand through the opening, you can retrieve the marble.

As we have seen, autists are in firm possession of the first rule, but fail to acquire the second rule, even when it is demonstrated to them by the experimenter. Thus, even though the experimenter exhibits a *temporal* sequence ‘throw switch – reach through opening – retrieve marble’, the autistic child does not code this temporal sequence into a *causal* sequence which can become entrenched. In itself there is of course nothing strange in a temporal sequence not being recoded as a causal sequence: the talking outside that I heard a while ago has nothing to do with my typing these words. The phenomenon becomes a bit stranger if it is taken into account that the experimenter demonstrates the use of the switch to the child: normal pragmatics would lead the child to expect some kind of causal connection, even though it may not yet know which.

One way to gloss Russell’s proposal cited above is to say that humans can normally adaptively exchange a rule for a new one by exploiting the defeasible character of rules, in the sense that if the action dictated by a rule does not seem to work, this is attributed to the presence of an unknown precondition of the action, which one then must proceed to find. If this is what is going on, then an informative logical analysis of the box task can be given.

One general feature of the type of reasoning that is going on in the box task has been dubbed ‘closed world reasoning’ in AI, which, roughly speaking, counsels us to consider all propositions to be false which we do not have reason to assume to be true. For example, since we have no grounds for assuming that a different article with the same content as this one will be published shortly, we assume there will be no such article. Closed world reasoning is of special importance when reasoning with unknown preconditions of actions, as we will proceed to show by means of a more formal analysis of the box task.

The main premiss can be formulated as

- (1) If you reach for the marble through the opening *and there is nothing funny going on*, you can retrieve the marble.

where the italicized conjunct is the variable, assumed to be present always, for an unknown precondition. This conjunct occasions closed world reasoning of the form

- (2) I haven't *seen* anything funny.  
:: There *is* nothing funny going on.

Backward chaining then leads to the plan

- (3) To get the marble, put your hand through the opening.

Now a problem occurs: the marble drops out of reach before it can be retrieved. Premiss (1) is now used to derive

- (4) Something funny is going on.

To determine what's so funny, the information about the switch is recruited, which can be formulated as a rule 'repairing' (1) as in (5-a) or (5-b)

- (5) a. If you set the switch to the right position *and there is nothing funny going on*, then you can retrieve the marble.  
b. If the switch is in the wrong position, there is something funny going on.

Closed world reasoning with (5-b) now yields

- (6) If the switch is in the wrong position, there is something funny going on, *but only then*.

Backward chaining then leads to a new plan

- (7) To get the marble, put your hand through the opening *and* set the switch to the right position.

One interesting feature of this analysis is thus that the new plan (7) is constructed from the old one by utilizing the variable for the unknown precondition. This is our proposed formalization of flexibility in planning; concomitantly, inflexibility would be characterized by failure to utilize that variable. After this logical analysis, we can now finally come to the point: we hypothesize that autists have difficulty with the reasoning *pattern* outlined here, in particular with applying closed world reasoning to unknown preconditions of actions, here formulated as the condition *there is nothing funny going on*. How can one test this? Interestingly, an experiment by Byrne [1] does precisely this, at least when looked at from the right angle.

## 4 The ‘suppression effect’

Suppose one presents a subject with the following innocuous premisses:

- (8) *If Marian has an essay to write she will study late in the library.  
Marian has an essay to write.*

In this case roughly 95% of subjects draw the conclusion ‘She will study late in the library’. We later return to the question what the remaining 5% may be thinking. Next suppose one adds the premiss

- (9) *If the library is open, Marian will study late in the library.*

In this case, only 60% concludes ‘She will study late in the library’.

However, if instead of (9) the premiss

- (10) *If Marian has a textbook to read, she will study late in the library.*

is added, then the percentage of ‘She will study late in the library’-conclusions is again 95%.

These observations are originally due to Ruth Byrne [1] (see also Dieussaert et al.[3]), and they were used by her to argue against a rule-based account of logical reasoning such as found in, e.g., Rips [11]. For if valid arguments can be suppressed, then surely logical inference cannot be a matter of blindly applying rules; and furthermore the fact that suppression depends on the *content* of the added premiss is taken to be an argument against the role of logical *form* in reasoning. We believe that this type of argumentation is somewhat off the mark (see Stenning and van Lambalgen [15, 13]), but here we concentrate on the relevance of this interesting experimental paradigm for the study of reasoning deficits in autism.

Byrne investigated not only *modus ponens* (MP), but also *modus tollens* (MT), and the fallacies (in classical logic) *affirmation of the consequent* (AC), and *denial of the antecedent* (DA), with respect to both types of added premisses, (9) and (10). Note that in the case of premiss type (9), the second conditional premiss highlights a presupposition that is at most implicit in the first conditional premiss: one can study in a library only if it is open. This is like the ‘repair’ conditions in our discussion of the box task. If on the other hand, the second conditional premiss is of type (10) there is no such relation. This premiss only provides another possible reason for studying in the library, not a precondition for this activity. This type of premiss will be important in the discussion of the fallacies.

We will not discuss all possible combinations of argument pattern and premiss type (see [13] for this) but we highlight the cases which are relevant to our discussion of autism.

**MT, premiss (9)** If Marian has an essay to write she will study late in the library.

If the library stays open then Marian will study late in the library.

Marian will not study late in the library.

In Byrne's experiment, 44% of the subjects concludes 'She does not have an essay to write', compared to 70% in the two-premiss case (8) – a clear case of suppression.

**AC, premiss (10)** If Marian has an essay to write she will study late in the library.

If Marian has some textbooks to read, she will study late in the library.

Marian studies late in the library.

Now 16% responds 'She has an essay to write', compared to 55% in (8); hence also fallacies can be suppressed.

**DA, premiss (10)** If Marian has an essay to write she will study late in the library.

If Marian has some textbooks to read, she will study late in the library.

Marian does not have an essay to write.

22% concludes 'She will not study late in the library', compared to 50% in (8) – again a clear case of the suppression of a fallacy.

**MT, premiss (10)** If Marian has an essay to write she will study late in the library.

If Marian has some textbooks to read, she will study late in the library.

Marian will not study late in the library.

70% concludes 'She does not have an essay to write', the same percentage as in the two-premiss case (8).

#### 4.1 A formal analysis

Byrne viewed the suppression effect mainly as showing that subjects are not guided by the rules of classical logic, but instead let their inferences

be determined by semantic content. We believe a more informative account of the suppression effect can be given, also establishing its relevance outside the reasoning domain, namely as showing that (normal) subjects are capable of flexible management of rules in context. For instance, normal subjects generally allow rules to have exceptions (and actions to have unknown preconditions), and they are quite good at exception-handling. This capacity involves some form of closed world reasoning, which counsels to take an exception into account if and only if one is forced to do so. To take our paradigmatic example, in

‘If Marian has an essay to write she will study late in the library.  
Marian has an essay to write.’

no exception is made salient, therefore the subject can draw the *modus ponens* inference: ‘She will study late in the library’. The addition of premiss (9)

‘If the library is open, Marian will study late in the library’

makes salient a possibly disabling condition in first rule, namely the library’s being shut. But since no other disabling conditions are mentioned, it is assumed that there aren’t any. The task at hand is to turn this intuition into a formal model.

Here we sketch how a formal analysis could go; the full technical treatment can be found in Stenning and van Lambalgen [13]. Speaking informally, we represent conditionals such as

If Marian has an essay to write she will study late in the library.

as *defaults* of the form

If Marian has an essay to write, *and nothing abnormal is the case*, she will study late in the library.

As in our attempted formalization of the box task, the italicized phrase introduces an overt marker for a possible abnormality or unknown precondition, which can be given concrete semantic content by other material given by the discourse. The claim is that this is a natural thing for a subject to do, because most rules indeed have exceptions, or unstated preconditions.

For readers not familiar with logical notation, we first explain our choice of symbols:  $\wedge$  means ‘and’,  $\neg$  means ‘not’,  $\vee$  means ‘or’,  $\rightarrow$  means ‘if ... then’ and  $\leftrightarrow$  means ‘if and only if’. We sometimes use  $\surd$

to mean an ‘or’ which takes more than two arguments. We will say very little about the logical properties of these connectives. Suffice it to say here that ‘and’, ‘or’, ‘not’ mean what you think they mean. The semantics of  $\rightarrow$  is somewhat involved, but here we use only the following property: from  $A \rightarrow B$  and  $A$  one may derive  $B$ .

Formally, we write a conditional as

$$p \wedge \neg ab \rightarrow q,$$

where  $ab$  is a proposition letter representing an unspecified abnormality. The logic governing  $ab$  is *closed world reasoning*. We will forego the general definition of this type of non-monotonic reasoning (including the properties of the implication symbol  $\rightarrow$  used in the formalization), but will illustrate the idea by means of examples related to the suppression task.

In general, one may give  $ab$  concrete content by adding implications of the form

$$s \rightarrow ab,$$

which express that the eventuality denoted by  $s$  constitutes an abnormality. Now suppose that there are  $n$  such implications in all, i.e., we have the implications

$$s_1 \rightarrow ab, \dots, s_n \rightarrow ab.$$

In the absence of further implications beyond the  $n$  mentioned, we want to conclude that we have listed *all* abnormalities. This can be done by *defining*  $ab$  as

$$ab \leftrightarrow \bigvee_{i \leq n} s_i.$$

Two special cases are of particular interest. If  $n = 1$ , i.e. if we only have the implication  $s \rightarrow ab$ , the definition yields  $ab \leftrightarrow s$ . Furthermore, for the case  $n = 0$ , the definition entails that  $ab$  is false, i.e.  $\neg ab$ . That is, if there is no information about the abnormality  $ab$ , we assume it does not occur.

These formal stipulations will help us explain the logic behind the suppression task. We do two illustrative cases; for the full treatment we refer to [13].

**Modus ponens** Consider again

If Marian has an essay to write she will study late in the library.

Marian has an essay to write.

Formally, this becomes

$$p \wedge \neg ab \rightarrow q; p.$$

Closed world reasoning yields  $\neg ab$ , which suffices to draw the conclusion  $q$ . Therefore modus ponens also follows in this nonclassical context, once closed world reasoning is applied. Failure to apply modus ponens may then be evidence of a resistance to apply closed world reasoning to the abnormality.<sup>3</sup>

The situation becomes slightly more complicated in the case of a further type (10) premiss:

If Marian has an essay to write she will study late in the library.

If Marian has an exam she studies late in the library.

Marian has an essay to write.

There are now two conditional premisses, each with its own disabling abnormality. The formalization thus becomes

$$p \wedge \neg ab \rightarrow q; r \wedge \neg ab' \rightarrow q; p.$$

Since the discourse does not provide information either about  $ab$  or about  $ab'$ , they are both set to false, that is, we have  $\neg ab$  and  $\neg ab'$ . The discourse thus becomes equivalent to

$$p \vee r \rightarrow q; p,$$

which again justifies the conclusion  $q$ .

Real complications arise in the case of a premiss of type (9):

If Marian has an essay to write she will study late in the library.

If the library is open Marian studies late in the library.

Marian has an essay to write.

Again there are two conditional premisses, each with its own disabling abnormality, but in this case there is interaction, because the antecedent of the second conditional highlights a possible precondition. The formalization is therefore not

$$p \wedge \neg ab \rightarrow q; r \wedge \neg ab' \rightarrow q; p,$$

as it was in the previous case, but rather

$$p \wedge \neg ab \rightarrow q; r \wedge \neg ab' \rightarrow q; \underline{\neg r \rightarrow ab}; p,$$

where the added underlined implication reflects the assumption that the second conditional has made an abnormality for the first conditional salient. Closed world reasoning applied to this implication yields  $ab \leftrightarrow \neg r$ , and if we then substitute  $r$  for  $\neg ab$  in the first conditional we get

$$p \wedge r \rightarrow q,$$

to which modus ponens can no longer be applied. The conclusion from this formal exercise is that suppression of modus ponens can be explained as an instance of closed world reasoning. This is definitely *not* to say that subjects *should* choose this underlying formal representation. It is very well possible to stick to the classical interpretation of the conditional, not containing a marker for a possible exception, in which case modus ponens should not be suppressed – indeed this is a plausible hypothesis to explain what autists appear to be doing.

**Denial of the antecedent** Fallacies and their suppression can be explained similarly. As an example we treat denial of the antecedent, in the case of the premisses

If Marian has an essay to write she will study late in the library.  
Marian does not have an essay to write.

The premisses can be formalized as

$$p \wedge \neg ab \rightarrow q; \neg p,$$

and since there is no information about  $ab$ , by closed world reasoning one may assume  $\neg ab$ . This particular fallacy involves more closed world reasoning however: one also has to assume that, in the absence of further information,  $p \wedge \neg ab$  is the *only* reason to conclude  $q$ , so that we have in effect

$$q \leftrightarrow p \wedge \neg ab.$$

Given  $\neg p$ , it indeed follows from this that  $\neg q$ .

Suppose we now add a further conditional premiss of type (10), to get

If Marian has an essay to write she will study late in the library.  
If Marian has an exam she studies late in the library.  
Marian does not have an essay to write.

The formalization is

$$p \wedge \neg ab \rightarrow q; r \wedge \neg ab' \rightarrow q; \neg p.$$

Closed world reasoning yields  $\neg ab$  and  $\neg ab'$ , which reduces the formalized premisses to

$$p \rightarrow q; r \rightarrow q; \neg p.$$

Closed world reasoning applied to the two implications  $p \rightarrow q$  and  $r \rightarrow q$  yields

$$q \leftrightarrow p \vee r,$$

from which given only  $\neg p$  nothing follows. The addition of the second conditional premiss may thus lead to a suppression of DA inferences.

It is of some importance for our discussion of the autism data to distinguish the two forms of closed world reasoning that play a role here. On the one hand there is the closed world reasoning applied to abnormalities or exceptions, which takes the form: ‘assume only those exceptions occur which are explicitly listed’. On the other hand there is the closed world reasoning applied to rules, which takes the form of diagnostic reasoning: ‘if  $B$  has occurred and the only known rules with  $B$  as consequent are  $A_1 \rightarrow B, \dots, A_n \rightarrow B$ , then assume one of  $A_1, \dots, A_n$  has occurred’. These forms of closed world reasoning are in principle independent, and in our autistic population we indeed see a dissociation between the two.

## 5 Autists’ performance in the suppression task

Given the formal analogy between the box task and the suppression task, we are led to expect that autists have a very specific difficulty with closed world reasoning about exceptions. This should show up in a refusal to suppress the inferences MP and MT in case the second conditional premiss is of the form (9). To show that the problem is really specific to exceptions, and not a more general problem about integrating new information, one may compare autists’ reasoning with AC and DA, in which case suppression is independent of exception-handling. Here one would expect behaviour which is comparable to normals.

In order to test these hypotheses, formulated generally as

autism is characterized by decreased ability in handling exceptions to rules

we conducted an experiment on a population of 6 autists with normal intelligence and language abilities from a psychiatric hospital in Vught (Netherlands). The tests administered to the subjects involved a false

belief task (the ‘Smarties’ task) propositional reasoning with 2 premisses (MP, MT etc.), the Wason selection task, the suppression task, reasoning with prototypes, and analogical reasoning. The method consisted in having tutorial interviews with the subjects, which were taped and transcribed (including annotation for pauses and emphases).<sup>4</sup> A full analysis will appear in the second author’s MSc thesis; here we concentrate on the suppression effect.

The results were quite striking. It is true that the small numbers involved do not allow one to draw statistically significant conclusions; on the other hand, the determination with which subjects resist the experimenter’s suggestions, which is only visible in this type of data, gives some reason to trust in the robustness of the results. Be that as it may, in this condition (the ‘library’ sentences) all 6 subjects refused to suppress in the case of MP. MT was considered to be more difficult even in the case of one conditional premiss, but the 4 subjects who applied MT there suppressed it for two conditional premisses. The 4 subjects who applied AC in the case of one conditional premiss, suppressed these inferences in the case of two conditional premisses. Of the 5 subjects who applied DA in the case of one conditional premiss, 3 suppressed these inferences in the case of two conditional premiss, and 2 didn’t. What is of especial relevance is that the experimenter’s interventions, pointing to the possible relevance of the second conditional premiss, had no effect!

We now present some conversations with our subjects while engaged in the suppression task. The subjects were presented with either two or three premisses, and were asked whether another sentence was true, false or undecided. We then proceeded to ask them for a justification of their answer.

**Excerpts from dialogues: MP** We recall the argument:

If Marian has an essay to write she will study late in the library.  
 (\*) If the library stays open then Marian will study late in the library.  
 Marian has an essay to write.  
 Does Marian study late in the library?

Here is subject C, first engaged in the two-premiss case, i.e. without (\*):

C: But that’s what it says!  
 E: What?  
 C: If she has an essay then she studies.  
 E: So your answer is ‘yes’?  
 C: Yes.

. The same subject engaged in the three-premiss argument:

C. Yes, she studies late in the library.  
E. Ehh, why?  
C. Because she *has to write* an essay.

Clearly the possible exception highlighted by the second conditional is not integrated; the emphasis shows that the first conditional completely overrides the second.

**Excerpts from dialogues: MT** In this case the argument pattern is

If Marian has an essay to write she will study late in the library.  
(\*). If the library stays open then Marian will study late in the library.  
Marian will not study late in the library.  
Does Marian have an essay?

Here is again subject C:

C. No, she has . . . oh no, wait a minute . . . this is a bit strange isn't it?  
E. Why?  
C. Well, it says here: if she *has to write* an essay . . . And I'm asked whether she has to write an essay?  
E. Mmm.  
C. I don't think so.

This is probably evidence of the general difficulty of MT, but note that the second conditional does not enter into the deliberations. In the dialogue, E. then prompts C. to look at second conditional, but this has no effect: C. sticks to his choice.

Here is a good example of the way in which a subject (in this case B) can be impervious to the suggestions of the experimenter. The dialogue refers to the argument with three premisses; we give a rather long abstract to show the flavour of the conversations.

B: No. Because if she had to make an essay, she would study in the library.  
E: Hmhm.  
B: And she doesn't do this, so she doesn't have an essay.  
E: Yes.  
B: And this means . . . (inaudible)  
E: (laughs) But suppose she has an essay, but the library is closed?  
B: Ah, that's also possible.  
E: Well, I'm only asking.  
B: Well, according to these two sentences that's not possible, I think.  
E: How do you mean? B: Ehm, yes she just studies late in the library if she has an essay.

E: Hmhm.  
 B: And it does not say ‘if it’s open, or closed ...’  
 E: OK.  
 B: So according to these sentences, I know it sounds weird, but  
 ....  
 E: Yes, I  
 B: I know it sounds rather autistic what I’m saying now (laughs).  
 E: (laughs) B: Eh yes.  
 E: So it is like you said? Or perhaps that she  
 B: Yes, perhaps the library closes earlier?  
 E: You may say what you want! You don’t have to try to think of  
 what should be the correct answer!  
 B: OK, no, then I’ll stick to my first answer.  
 E: OK, yes.  
 B: (laughs) I know it’s not like this, but (laughs).  
 E: Well, that’s not clear. It’s possible to say different things about  
 reasoning here, and what you say is certainly not incorrect.

In the above we have seen examples of how our autistic subjects refuse to integrate the information about exceptions provided by (\*). The next extracts show that this need not be because they are incapable of integrating a second conditional premiss, or of applying closed world reasoning. We consider the ‘fallacies’ DA and AC, which can be valid if seen as a consequence of closed world reasoning, and which can be suppressed by supplying a suitable second conditional premiss, e.g. (†) below.

**Excerpts from dialogues: AC** The argument is

If Marian has an essay to write she will study late in the library.  
 (†) If Marian has an exam, she will study late in the library.  
 Marian studies late in the library.  
 Does Marian have an essay?

Here is subject C, in the two-premiss argument without (†).

C: Yes.  
 E: Why?  
 C: It says in the first sentence ‘if she has an essay then she does that [study late etc.] ... But Marian is just a name, it might as well be Fred.

Now consider the three-premiss case.

C: Mmm. Again ‘if’, isn’t it?  
 E: Yes.  
 C: If Marian has an essay, she studies late in the library’...  
 E: Yes.

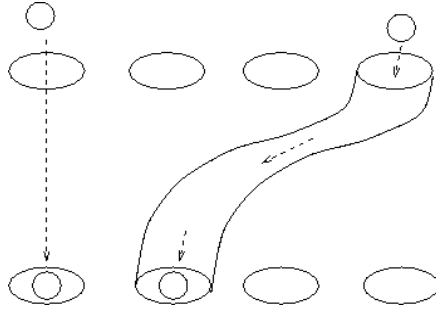


Figure 2: Russell’s tubes task

C: If Marian has an essay, she studies late in the library . . .  
 E: Hmhm.  
 C: Marian studies late in the library. Does Marian have an essay?  
 E: Hmhm.  
 C: No.  
 E: Hmhm.  
 C: It does say she has to make an essay.  
 E: Hmhm. But she studies late in the library, can you conclude from this that she has to make an essay?  
 C: No you can’t, because she could also have an exam.

We see in this example that C correctly judges the import of (†): after having applied closed world reasoning to the two-premiss case, he notices that it is powerless in this case.

## 6 Wrinkles

Here we discuss some potential problems with the ‘logical’ account of executive dysfunction raised by other pieces of data.

**A similar task with very different outcomes** The box task is superficially similar to another task devised by Russell, the ‘tubes task’.

What one sees in this schematic drawing is a series of four holes into which a ball can be dropped, to land in a small container below. A ball dropped through the leftmost opening will end up in the catch-tray directly underneath, but a ball dropped through the rightmost opening travels through an opaque tube to end up in the catch-tray which

is second from left. The child sees the ball being dropped through an opening, and has to retrieve it from one of the catch-trays below. When the ball is dropped in the rightmost opening, children of age 3 or younger tend to look in the catch-tray directly underneath the opening, probably applying the (defeasible) rule that things fall vertically. Older children, *including also autistic children*, manage to inhibit the ‘prepotent’ response and search in the correct catch-tray, adequately representing the trajectory of the ball as guided by the tube.

The apparent puzzle posed by performance on this task is that in this case autistic children *are* able to switch rules effortlessly. Russell explains this by a distinction between ‘arbitrary’ rules imposed by the experimenter (as in the box task), and rules based on fairly transparent physical principles (as in the tubes task). Autists would be impaired on the former but not the latter, incidentally showing that autism, viewed as executive dysfunction, must be a rather specific executive deficit. If both kinds of defeasible rules require the same kind of closed world reasoning about abnormalities, the hypothesis that it is this form of reasoning that is difficult for autists, is defeated.

The first thing to observe here is that the rules involved in the two tasks have different logical forms, and so require different reasoning. In the box task, correct performance hinges on the ability to amend the *antecedent* of the rule, whereas in the case of the tubes task it is the consequent (i.e. the catch-tray) that has to be changed. In the box task, the original plan has to be changed by incorporating another action,, whereas in the case of the tubes task one action has to be replaced by another. This suggests that what happens in the tubes task need not be viewed as rule-switching, but can also be seen as the application of a *single* IF-THEN-ELSE rule, where the action to be taken depends on the satisfaction or non-satisfaction of an explicit precondition: unimpeded fall of the ball. On this analysis, the difference between box task and tubes task would be that in the former case a new rule has to be synthesized on the spot by exception-handling, whereas in the latter case the switch is between components of a given single rule.

It seems that autistic subjects have less difficulty with synthesizing an IF-THEN-ELSE rule from instructions shown to them, than with rule-construction by exception-handling. Indeed, a standard ‘Go/No Go’ task is of the IF-THEN-ELSE form. For instance, in one such task, subjects were shown different letters of the alphabet that flashed one at a time on a computer screen. They were asked to respond by pressing a key in every case except when they saw the letter X. The first task was a Go task, in which the letter X never appeared and in this way subjects were allowed to build up a tendency to respond. Immediately afterward,

subjects performed a Go/No Go task in which the letter X did appear in the lineup, at which point the subject had to control the previously built impulse to respond. Autists are not particularly impaired at such a task, although they do become confused when they have to shift rapidly between one target stimulus and another (Ozonoff et al. [10]).

At a more abstract level, what the analysis of the empirical difference between the box and tubes tasks just given highlights is that, before one can discuss whether autists have difficulties with rule-switching, the proper definition of ‘rule’ in this context has to be clarified. If the preceding considerations are correct, than a rule can be more general than the ‘condition – action’ format.

**Is it really only exception-handling?** It is in the spirit of this enterprise not to throw away data, and we therefore present some data from the dialogues which might indicate that the relation between autism and exception-handling is more complicated than the above dialogues taken by themselves would indicate. In particular there are some parts of dialogue which lead one to suspect that integration of the second conditional premiss in itself poses a difficult cognitive problem.

Here is subject A engaged in MT

If Marian has an essay to write she will study late in the library.

(\*) If the library stays open then Marian will study late in the library.

Marian will not study late in the library.

Does Marian have an essay?

A: It says here that if she has an essay, she studies late in the library... But she doesn't study late in the library ... which suggests that she doesn't have an essay ... But it could be that she has to leave early, or that something else came up ... *if you only look at the given data*, she will not write an essay ... but when you realize that something unexpected has come up, then it is very well possible that she does have to write an essay ... [our italics]

If this were the two-premiss argument, it would be near-perfect exception-handling: the italicized phrase doing duty as closed world reasoning applied to the two formalized premisses  $\{-q; p \wedge \neg ab \rightarrow q\}$ . But this subject refuses to apply MT in the two-premiss case, arguing that most likely something has come up which prevented Marian from studying late in the library.<sup>5</sup> Confusingly, the subject first appears to apply this reasoning in the three-premiss argument, then notices that something else may have come up, but does not *explicitly* relate this to (\*), and proceeds to reject MT here. (Since MT was not accepted in the two-premiss case,

this can not be described as suppression.) On the face of it, what happens here might as well be described as a case of failed integration of the second conditional premiss.

A more explicit failure to integrate can perhaps be observed in the following extract, pertaining to DA. Subject C has no hesitations in applying DA in the two-premiss case, and now proceeds with the three premisses

If Marian has an essay to write she will study late in the library.  
(†) If Marian has an exam, she will study late in the library.  
Marian does not have an essay.  
Does Marian study late in the library?

C: *No*, she doesn't study late in the library.  
E. Why not?

C: Because it says: '*if* she has an essay...' In any case, the exam could be left out, because we don't have any further information about it ...

Recall that we used performance on the 'fallacies' AC and DA to argue that autistic subjects are able to integrate information from a second conditional premiss, as long as it does not pertain to exceptions. Is the extract just given counter-evidence for this claim, because C does not suppress DA? This is not quite clear; it depends on how one reads the phrase 'In any case, the exam is irrelevant, because we don't have any further information about it'. If it is read as: 'I won't consider the second conditional premiss', then it is evidently non-integration. If however it is read as a form of closed world reasoning: 'we don't have any information about the exam, so we assume it doesn't happen', then DA is indeed a valid form of closed world reasoning. The latter interpretation gets some plausibility from the continuation of the dialogue

C: Perhaps she needs books for her essay ... but then it says nothing about books here ... so the answer is *no*, not 'perhaps yes, perhaps no', but definitely no.

To sum up these considerations, on the basis of these data it cannot be completely excluded that some subjects experience problems with the integration of the second conditional premiss, even disregarding reasoning with exceptions.

## 7 Conclusion: back to the box

In the analysis presented here there is a continuity between rigidity in motor behaviour, lack of flexibility in planning, and insensitivity to possible exceptions to rules. This may seem problematic because a 'lower'

cognitive ability such as motor planning is connected to a ‘higher’ cognitive function such as reasoning. One may therefore object that there is at most an analogy, but not a common substratum.

The question is, however, whether reasoning is necessarily such a high level process. In Stenning and van Lambalgen [13] a neural model for the suppression task, i.e. reasoning with exceptions, is presented which turns reasoning by and large into an automatic process. Indeed, if one recalls what was said above, that reasoning with exceptions is fundamental to planning, then one can imagine that automating this process is very useful since it increases speed. The supposed distance between motor behaviour and reasoning may then after all not be so large, which is interesting from an evolutionary point of view. One of us has in fact argued that the same planning mechanism appears to be operating from motor planning all the way to discourse integration; see van Lambalgen and Hamm [17].

Also, inspection of the neural model may generate hypotheses on possible differences in neural architecture between autists and normals. The details of the neural model are too complex to be given here, but very roughly speaking the model for reasoning with exceptions proposed in [13] has twice as many nodes, and many more inhibitory connections than models for exception-less reasoning. Due to the characteristic brainwaves associated to inhibition, it may then be possible to detect differences in neural architecture between autists and normals by means of electroencephalographic techniques such as ERP.

## 8 Notes

† Corresponding author. We are grateful to the Netherlands Organization for Scientific Research (NWO) for support under grant 360-80-000. We thank Merel Egtberts for transcribing the interviews with our autistic subjects. The research reported here forms part of a collaborative project on reasoning with Keith Stenning, and will be more extensively reported in the forthcoming [14].

1. Investigated experimentally by the first author’s students David Wood and Marian Counihan.
2. Some have indeed argued that performance on the false belief task can be explained as a linguistic phenomenon; see for example [7].
3. Here we will not discuss the ‘backward’ inferences *modus tollens* and ‘affirmation of the consequent’, which require a slightly more subtle form of closed world reasoning. We want to note here, however, that failure to endorse *modus tollens* may similarly reflect a refusal to apply closed world reasoning to the abnormality.

4. For a defense of this kind of data see Stenning and van Lambalgen [15].
5. This subject is also quite hesitant about MP for one conditional premiss for analogous reasons, thus providing evidence for a representation of the conditional with exceptions built in, together with a reluctance to apply closed world reasoning to these exceptions.

## References

- [1] R.M.J. Byrne. Suppressing valid inferences with conditionals. *Cognition*, 31:61–83, 1989.
- [2] N. Chater and M. Oaksford. Yje probability heuristics model of syllogistic reasoning. *Cognitive Psychology*, 38:191–258, 1999.
- [3] K. Dieussaert, W. Schaeken, W. Schroyen, and G. d’Ydewalle. Strategies during complex conditional inferences. *Thinking and reasoning*, 6(2):125–161, 2000.
- [4] J. Gibson. *The ecological approach to visual perception*. Houghton-Mifflin Co., Boston, 1979.
- [5] C. Hughes and J. Russell. Autistic children’s difficulty with disengagement from an object: its implications for theories of autism. *Developmental Psychology*, 29:498–510, 1993.
- [6] W. Koehler. *The mentality of apes*. Harcourt Brace and World, New York, 1925.
- [7] C. Lewis and A. Osborne. Three-year olds problems with false belief: conceptual deficit or linguistic artefact? *Child Development*, 61:1514–1519, 1990.
- [8] B. McGonigle, M. Chalmers, and A. Dickinson. Concurrent disjoint and reciprocal classification by *cebus apella* in serial ordering tasks: evidence for hierarchical organization. *Animal Cognition*, In press.
- [9] F.T. Melges. Identity and temporal perspective. In R.A. Block, editor, *Cognitive models of psychological time*, pages 37–58. Lawrence Erlbaum, 1990.
- [10] S. Ozonoff, D.L. Strayer, W.M. McMahon, and F. Filloux. Executive function abilities in children with autism and tourette syndrom: an information-processing approach. *Journal of Child Psychology and Psychiatry*, 35:1015–1032, 1994.
- [11] L.J. Rips. Cognitive processes in propositional reasoning. *Psychological Review*, 90:38–71, 1983.
- [12] J. Russell. Cognitive theories of autism. In J.E. Harrison and A.M. Owen, editors, *Cognitive deficits in brain disorders*, pages 295 – 323. Dunitz, London, 2002.

- [13] K. Stenning and M. van Lambalgen. A working memory model of relations between interpretation and reasoning. 2003. Submitted to *Cognitive Science*.
- [14] K. Stenning and M. van Lambalgen. *Human reasoning and cognitive science*. MIT University Press, Cambridge, MA., 2004.
- [15] K. Stenning and M. van Lambalgen. A little logic goes a long way: basing experiment on semantic theory in the cognitive science of conditional reasoning. *Cognitive Science*, July 2004.
- [16] M. Tomasello. *Constructing a language. A usage-based theory of language acquisition*. Harvard University Press, Boston, 2003.
- [17] M. van Lambalgen and F. Hamm. *The proper treatment of events*. To appear with Blackwell Publishing, Oxford and Boston, 2004. Until publication, manuscript available at <http://staff.science.uva.nl/~michiell>.